

# Improving Human Text Comprehension through Semi-Markov CRF-based Neural Section Title Generation

Sebastian Gehrmann, Steven Layne, Franck Dernoncourt



@SebGehr



gehrmann@seas.harvard.edu



**Harvard** John A. Paulson  
**School of Engineering**  
and Applied Sciences



Adobe Research

# Long texts are hard to comprehend

When another old cave is discovered in the south of France, it is not usually news. Rather, it is an ordinary event. Such discoveries are so frequent these days that hardly anybody pays heed to them. However, when the Lascaux cave complex was discovered in 1940, the world was amazed. Painted directly on its walls were hundreds of scenes showing how people lived thousands of years ago. The scenes show people hunting animals, such as bison or wild cats. Other images depict birds and, most noticeably, horses, which appear in more than 300 wall images, by far outnumbering all other animals.

Early artists drawing these animals accomplished a monumental and difficult task. They did not limit themselves to the easily accessible walls but carried their painting materials to spaces that required climbing steep walls or crawling into narrow passages in the Lascaux complex. Unfortunately, the paintings have been exposed to the destructive action of water and temperature changes, which easily wear the images away. Because the Lascaux caves have many entrances, air movement has also damaged the images inside. Although they are not out in the open air, where natural light would have destroyed them long ago, many of the images have deteriorated and are barely recognizable.

To prevent further damage, the site was closed to tourists in 1963, 23 years after it was discovered.

# Can Summaries Help?

## **Lascaux cave complex discovered**

When another old cave is discovered in the south of France, it is not usually news. Rather, it is an ordinary event. Such discoveries are so frequent these days that hardly anybody pays heed to them. However, when the **Lascaux cave complex** was discovered in 1940, the world was amazed. Painted directly on its walls were hundreds of scenes showing how people lived thousands of years ago. The scenes show people hunting animals, such as bison or wild cats. Other images depict birds and, most noticeably, horses, which appear in more than 300 wall images, by far outnumbering all other animals.

## **Paintings exposed to destructive action**

Early artists drawing these animals accomplished a monumental and difficult task. They did not limit themselves to the easily accessible walls but carried their painting materials to spaces that required climbing steep walls or crawling into narrow passages in the Lascaux complex. Unfortunately, the **paintings** have been **exposed to the destructive action** of water and temperature changes, which easily wear the images away. Because the Lascaux caves have many entrances, air movement has also damaged the images inside. Although they are not out in the open air, where natural light would have destroyed them long ago, many of the images have deteriorated and are barely recognizable.

## **Site closed to tourists**

To prevent further damage, the **site** was closed to tourists in 1963, 23 years after it was discovered.

We hypothesize that summaries, presented as section titles, can improve the fact **retention**, fact **retrieval** and text **comprehension**.

## Goals

- (1) A low-resource approach to generating section titles
- (2) An evaluation framework for comprehension tasks

# Section Titles in Two Steps

## (1) Select the most important Sentence

$$\operatorname{argmax}_{\text{sent} \in \text{para}} \text{saliency}(\text{sent})$$

## (2) Compress the sentence

$$x_i \forall (x_1, y_1), \dots, (x_n, y_n) \text{ iff } y_i = 1$$

When another old cave is discovered in the south of France, it is not usually news. Rather, it is an ordinary event. Such discoveries are so frequent these days that hardly anybody pays heed to them. **However, when the Lascaux cave complex was discovered in 1940, the world was amazed.** Painted directly on its walls were hundreds of scenes showing how people lived thousands of years ago. The scenes show people hunting animals, such as bison or wild cats. Other images depict birds and, most noticeably, horses, which appear in more than 300 wall images, by far outnumbering all other animals.

However, when the Lascaux cave complex was discovered in 1940, the world was amazed.

Lascaux cave complex discovered

# Section Titles in Two Steps

## (1) Select the most important Sentence

$$\operatorname{argmax}_{\text{sent} \in \text{para}} \text{saliency}(\text{sent})$$

When another old cave is discovered in the south of France, it is not usually news. Rather, it is an ordinary event. Such discoveries are so frequent these days that hardly anybody pays heed to them. **However, when the Lascaux cave complex was discovered in 1940, the world was amazed.** Painted directly on its walls were hundreds of scenes showing how people lived thousands of years ago. The scenes show people hunting animals, such as bison or wild cats. Other images depict birds and, most noticeably, horses, which appear in more than 300 wall images, by far outnumbering all other animals.

However, when the Lascaux cave complex was discovered in 1940, the world was amazed.

## (2) Compress the sentence

$$x_i \forall (x_1, y_1), \dots, (x_n, y_n) \text{ iff } y_i = 1$$

Lascaux cave complex discovered

# Section Titles in Two Steps

## (1) Select the most important Sentence

$$\operatorname{argmax}_{\text{sent} \in \text{para}} \text{saliency}(\text{sent})$$

When another old cave is discovered in the south of France, it is not usually news. Rather, it is an ordinary event. Such discoveries are so frequent these days that hardly anybody pays heed to them. **However, when the Lascaux cave complex was discovered in 1940, the world was amazed.** Painted directly on its walls were hundreds of scenes showing how people lived thousands of years ago. The scenes show people hunting animals, such as bison or wild cats. Other images depict birds and, most noticeably, horses, which appear in more than 300 wall images, by far outnumbering all other animals.

However, when the Lascaux cave complex was discovered in 1940, the world was amazed.

## (2) Compress the sentence

$$x_i \forall (x_1, y_1), \dots, (x_n, y_n) \text{ iff } y_i = 1$$

Lascaux cave complex discovered

# Compressive Deletion

$x$  : However, when the Lascaux cave complex was discovered in 1940, the world was amazed.

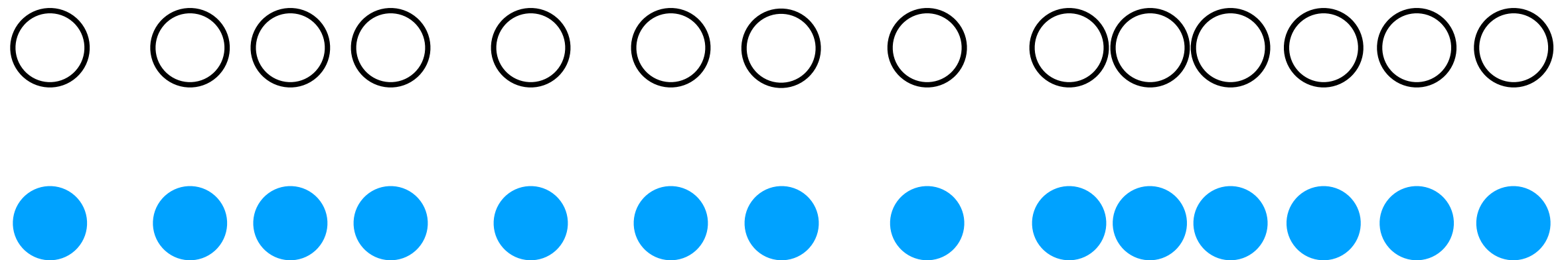


Lascaux cave complex      discovered

$y$  :    0        0    0        1        1        1        0        1        0    0        0        0        0        0

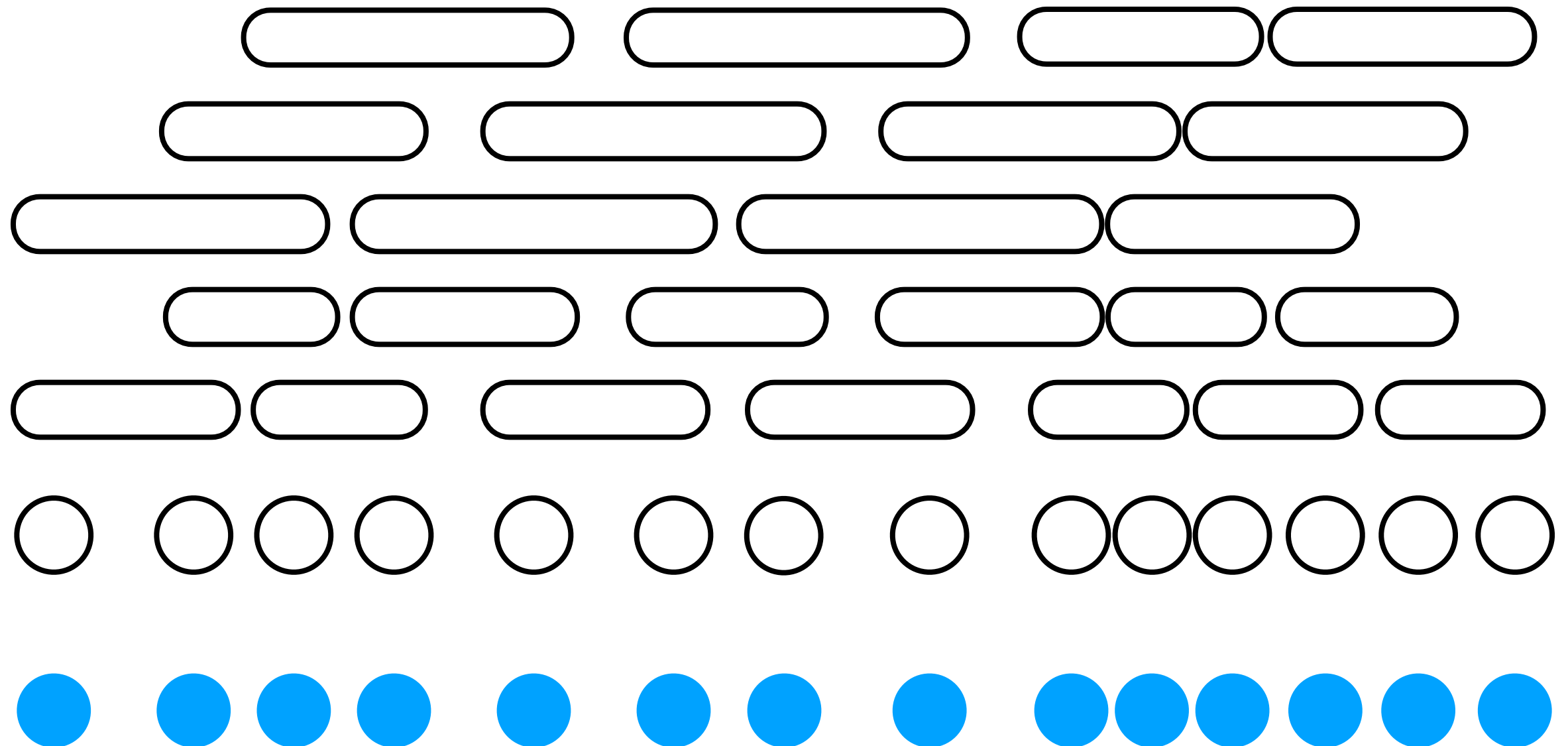


# Data-Efficiency through SCRFs



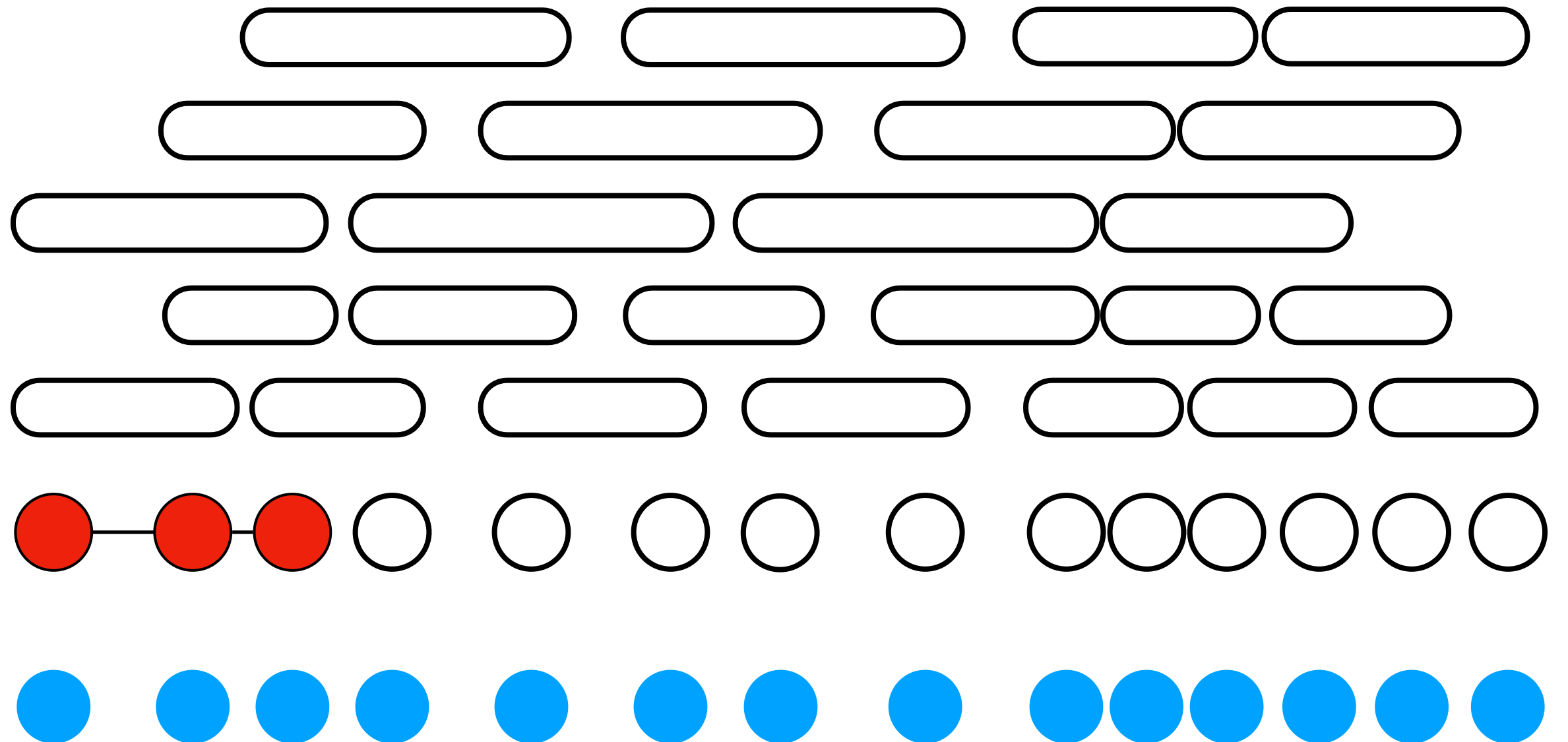
However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFs



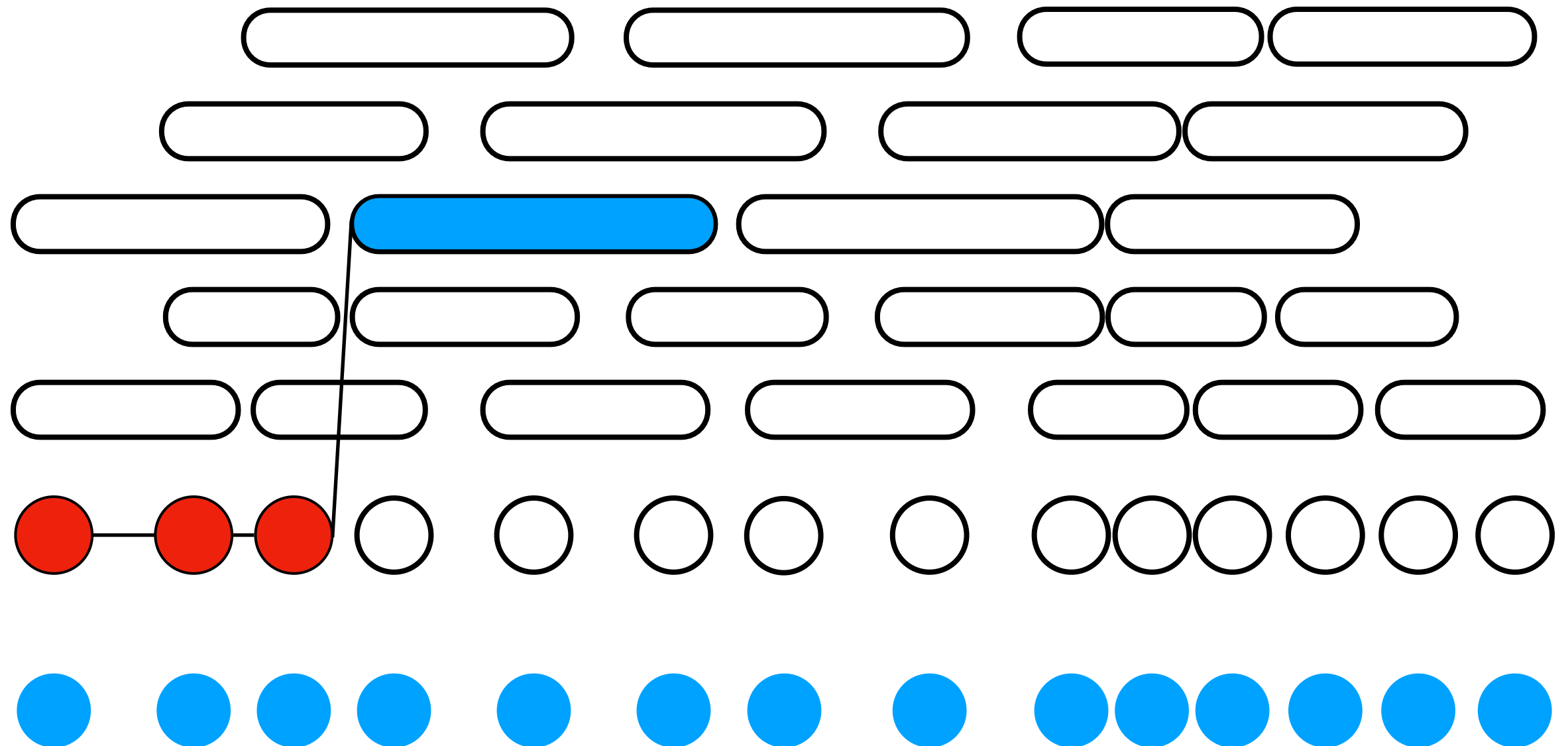
However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFs



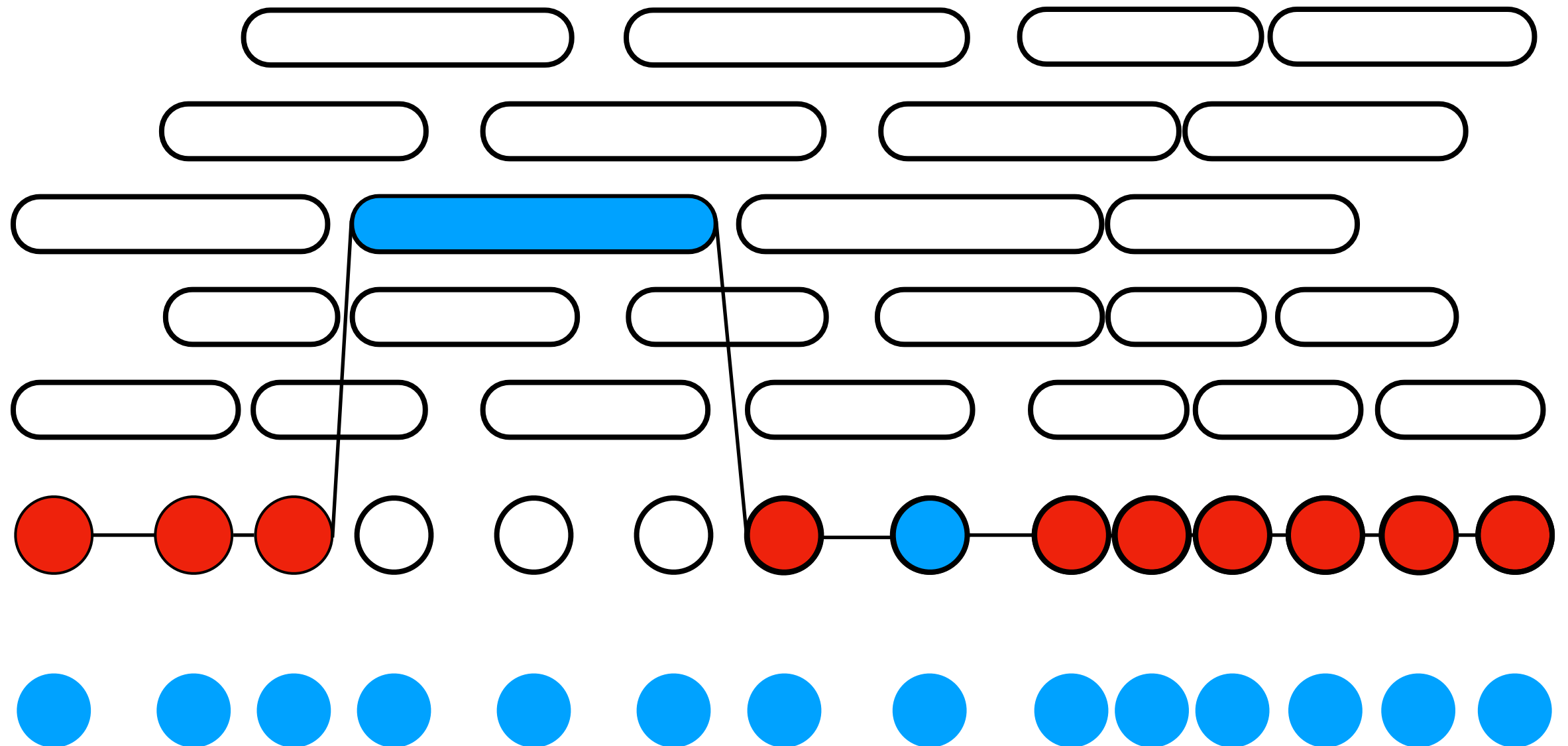
However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFs



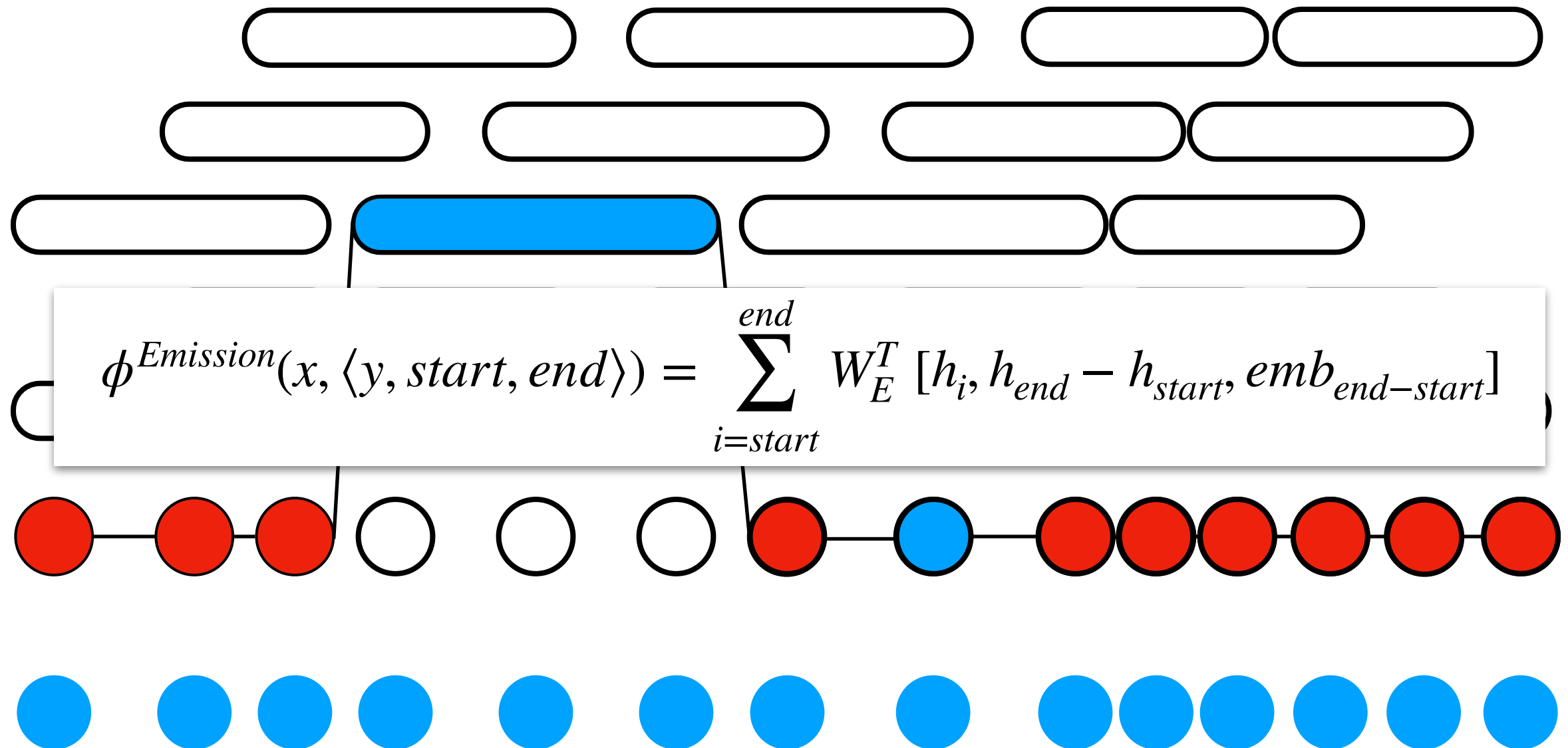
However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFs



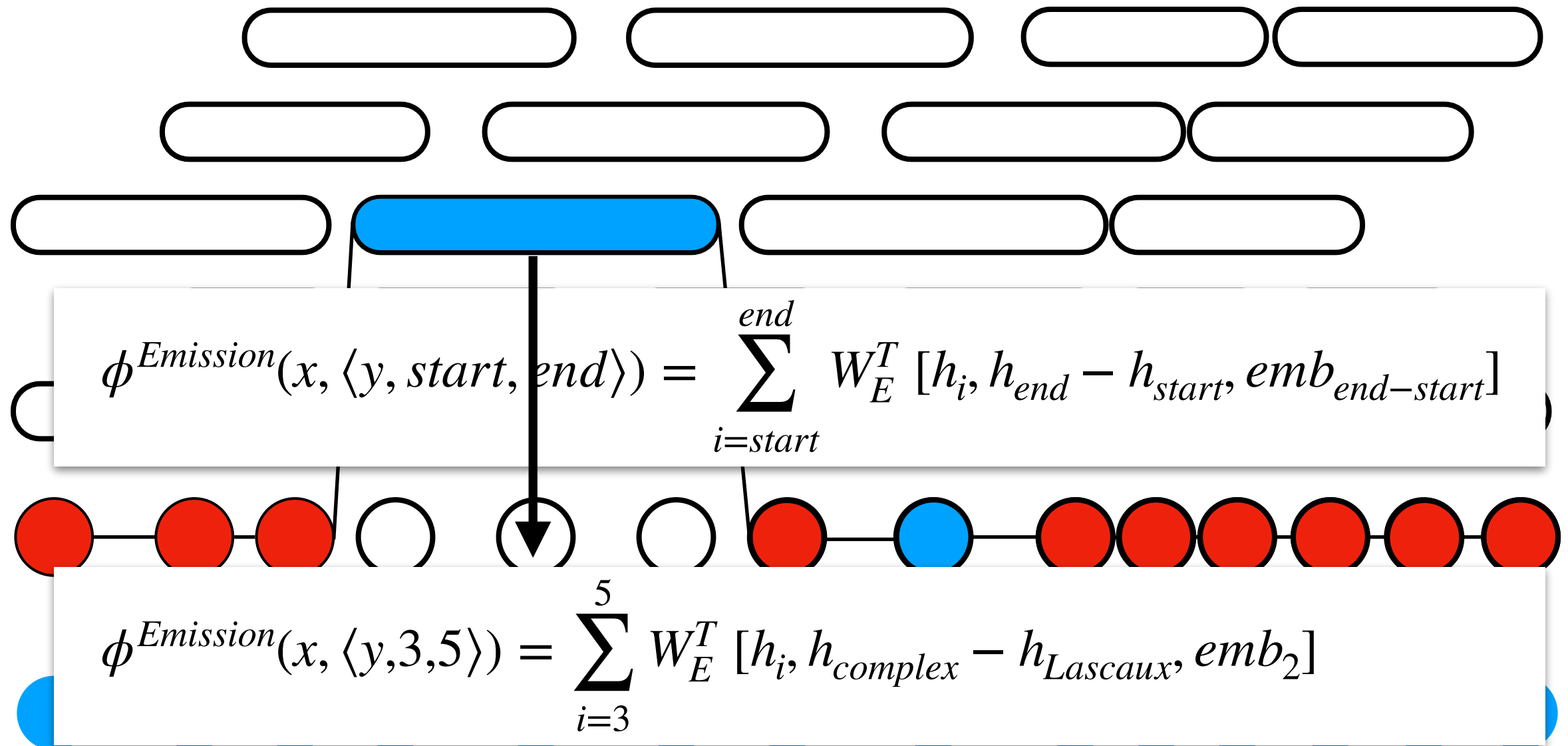
However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFS



However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

# Data-Efficiency through SCRFS



However, when the **Lascaux cave complex** was **discovered** in 1940, the world was amazed.

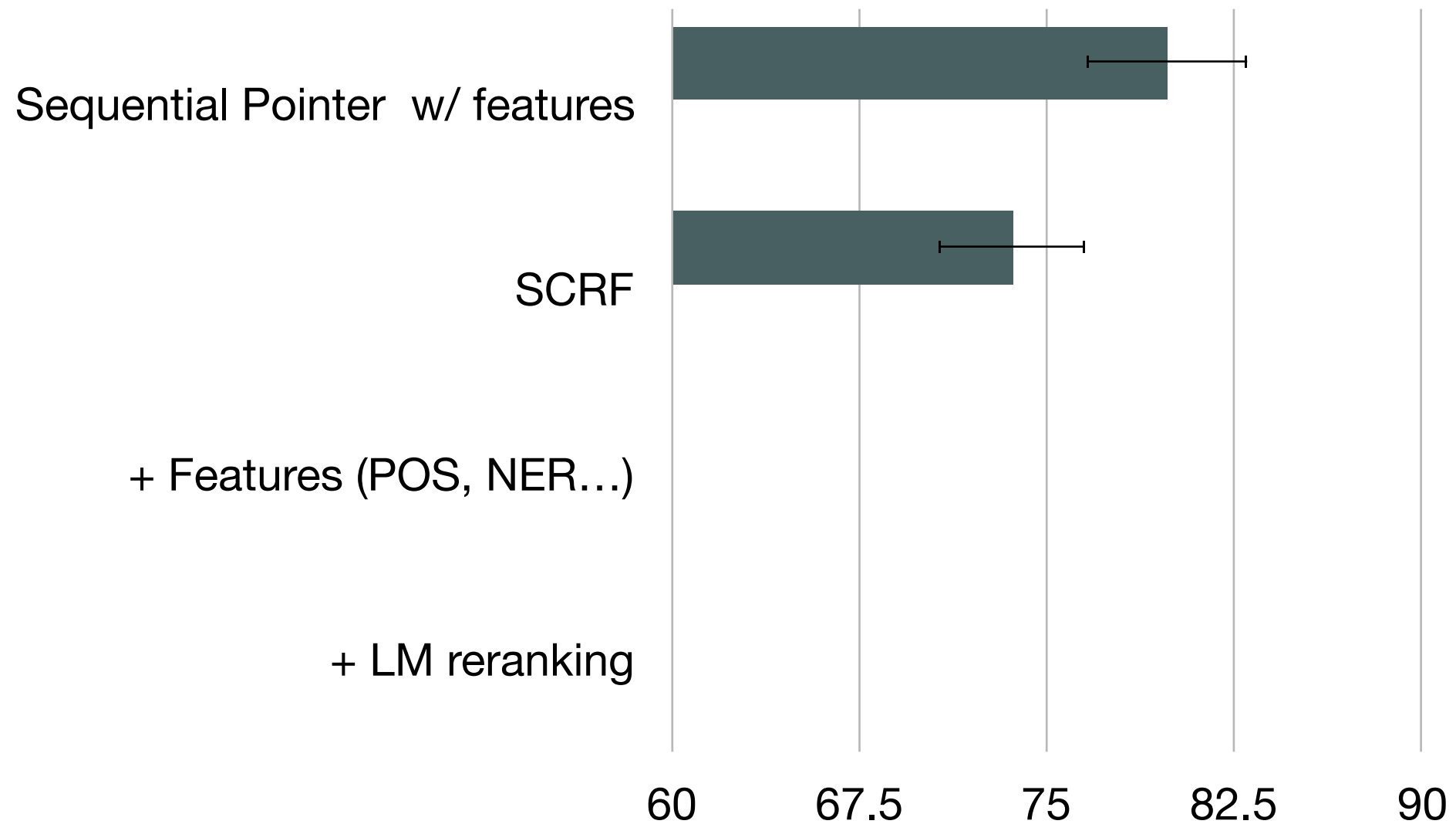
# Comparison to S2S

**On 200,000 data points**



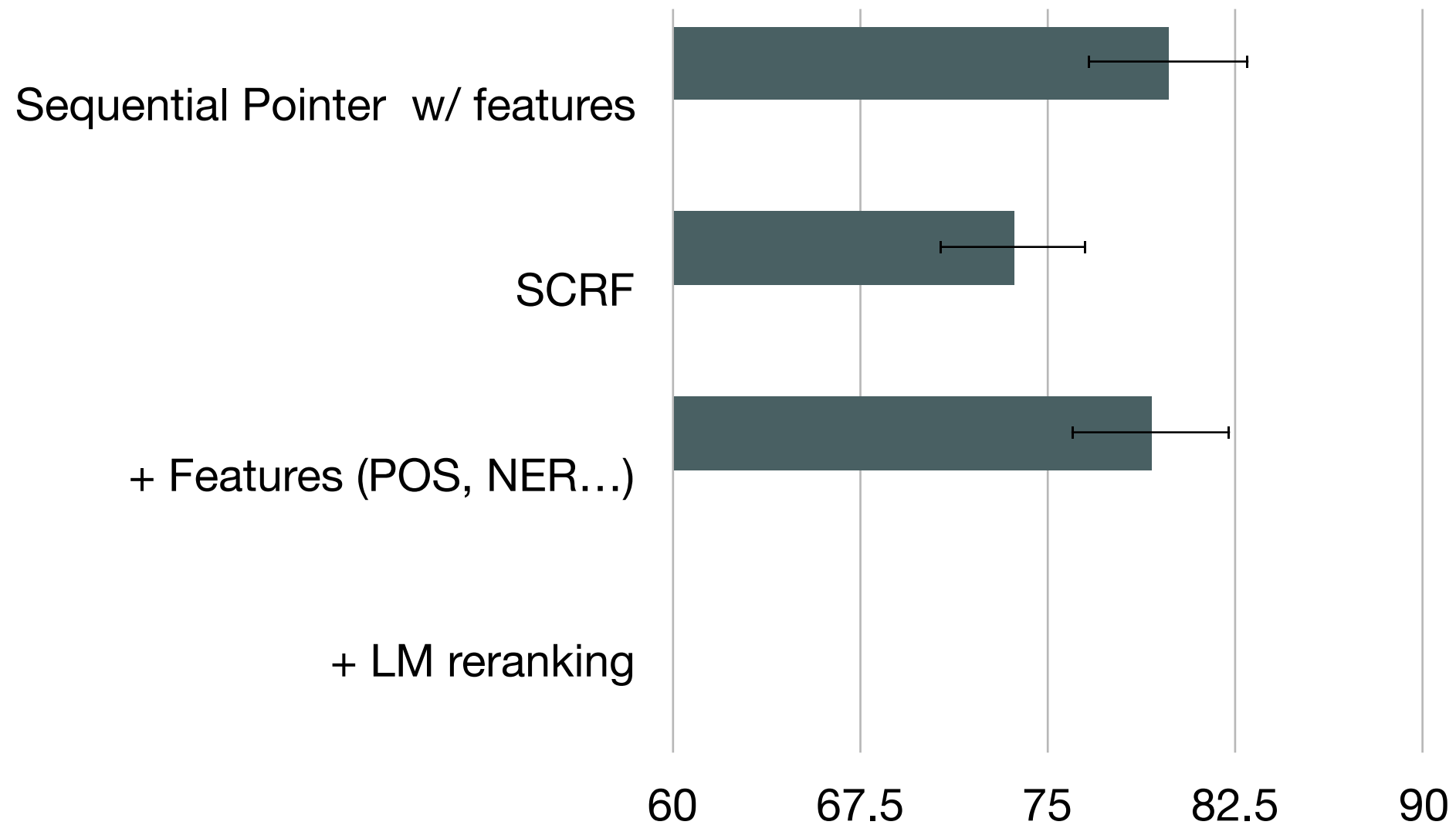
# Comparison to S2S

On 200,000 data points



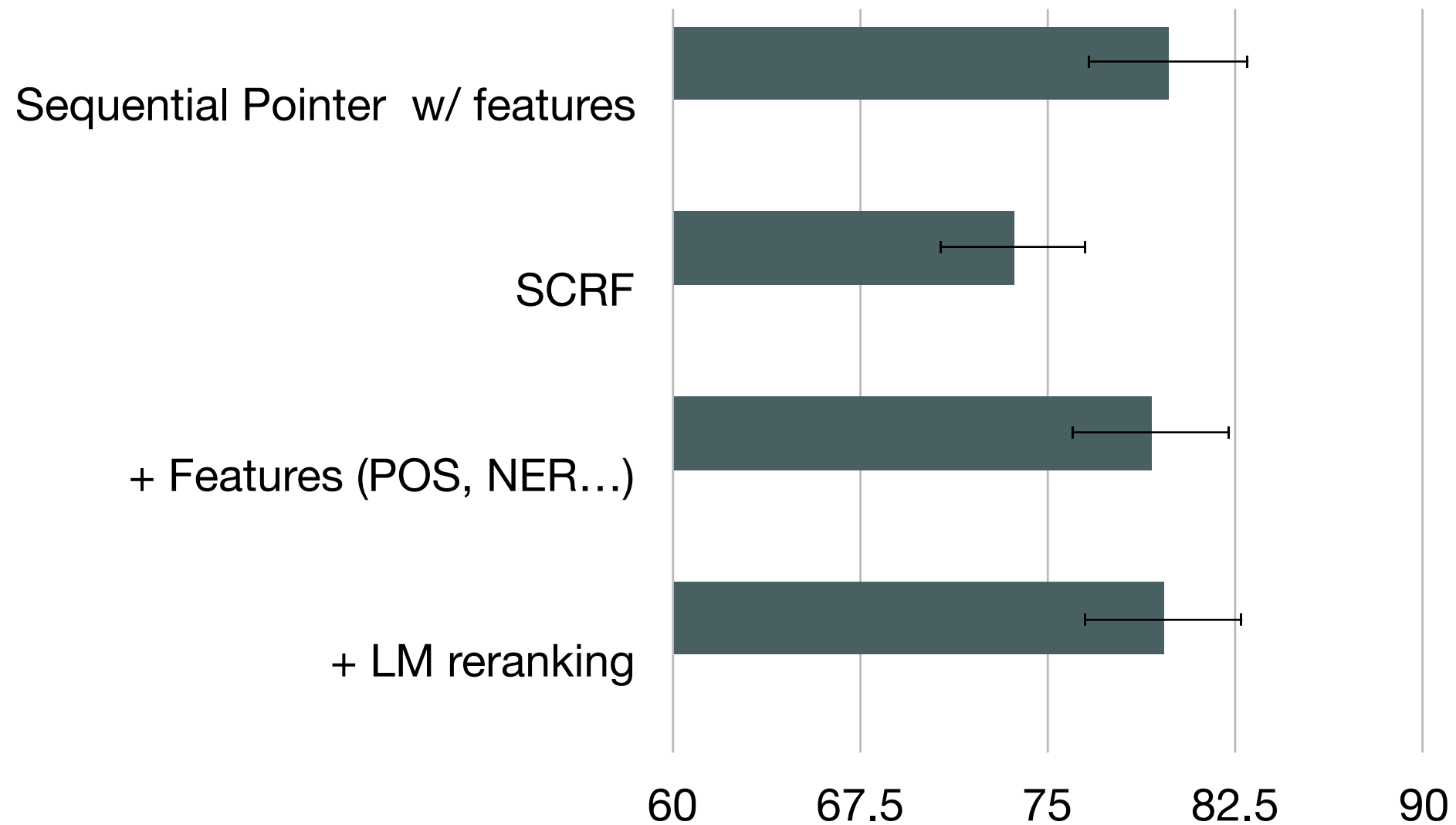
# Comparison to S2S

On 200,000 data points



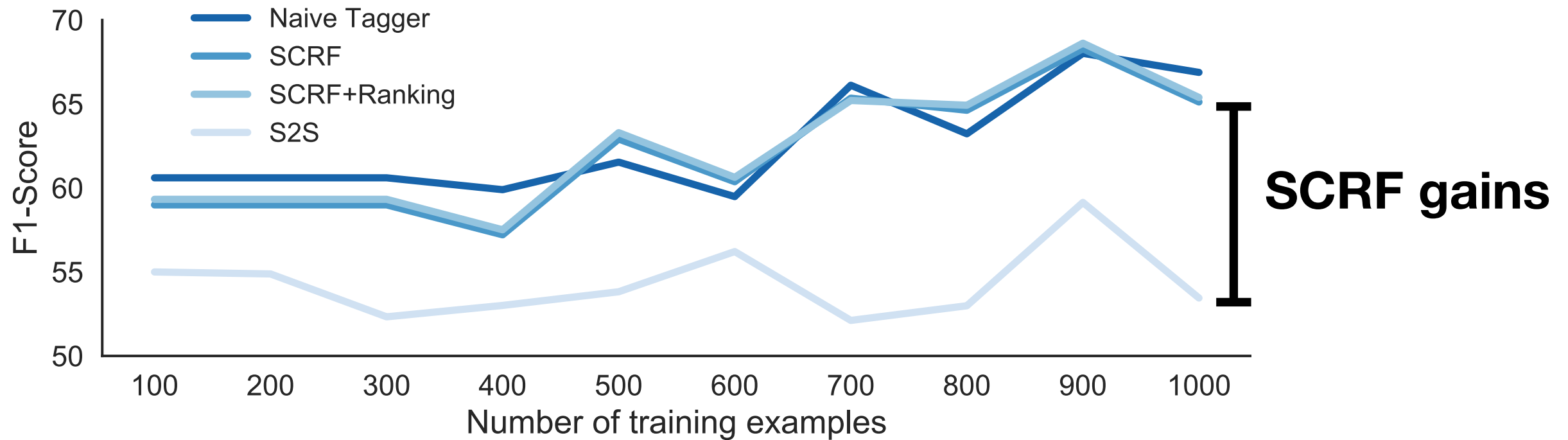
# Comparison to S2S

On 200,000 data points



# Comparison to S2S

On limited data



# But does it improve comprehension?

National Geographic interactive reading practice:  
33 texts, 4-7 paragraphs, two reading difficulties, various topics

We compare **no** titles, **human-written** titles, and **generated** titles

**French cities set up shelters.**

And yet, there are ways of dealing with the effects of such extreme events. After 2003, French cities set up air-conditioned shelters for use in heat waves. In the 2006 heat wave, the death rate was two-thirds lower.

**Warming is putting more moisture.**

'We know that warming of the Earth's surface is putting more moisture into the atmosphere. We've measured it. The satellites see it,' says climatologist Jay Gullede. Another scientist, Michael Oppenheimer, agrees. We need to face up to that reality, he says, and do the things we know can save lives and money.

**144 participants** completed six 2-3 min long tasks

**The influences are natural cycles in the climate.**

So, what is happening to our weather? Are these extreme events part of a natural cycle? Or are they caused by human activity and its effects on the Earth's climate? Peter Miller says it's probably a mixture of both of these things. On the one hand, the most important influences on weather events are natural cycles in the climate. Two of the most famous weather cycles, El Niño and La Niña, originate in the Pacific Ocean. The heat from the warm ocean rises high into the atmosphere and affects weather all around the world. On the other hand, the temperature of the Earth's oceans is slowly but steadily going up. And this is a result of human activity. We are producing greenhouse gases that trap heat in the Earth's atmosphere. This heat warms up the atmosphere, land and oceans. Warmer oceans produce more water vapour – think of heating a pan of water in your kitchen. Turn up the heat, it produces steam more quickly. Satellite data tells us that the water vapour in the atmosphere has gone up by four percent in 25 years. This warm, wet air turns into the rain, storms, hurricanes and typhoons that we are increasingly experiencing.

**We need to act to save lives.**

Climate scientist Michael Oppenheimer says that we need to face the reality of climate change. And we also need to act now to save lives and money in the future.

---

The article says extreme weather is ...

- becoming more common.
- not a natural occurrence.
- hard for scientists to understand..

Extreme weather can be caused by .

# Retention

## Ask people to answer questions after reading



**TASK 1/6:** Please read the questions. Then, try to find the answers to the questions within the text.

Thor Heyerdahl ...

- proved who had created the statues.
- was wrong about the origins of the statues.
- thought the statues had been brought to the island from South America.

Jared Diamond thinks that ...

- the forest resources on Easter Island were poor.

**They were carved with stone tools.**

On a winter night last June, José Antonio Tuki, a 30-year-old artist on Easter Island, did one of the things he loves best: he left his house on the southwest coast and hiked north across the island to Anakena beach. Tuki sat on the sand and gazed at the colossal human statues – the moai. The statues range from four to 33 feet and some weigh more than 80 tons. They were carved long ago with stone tools and then transported up to 11 miles to massive stone platforms. As Tuki stares at their faces, he feels a connection. ‘It’s something strange and energetic,’ he says. ‘This is something produced from my culture, the Rapanui.’ He shakes his head. ‘How did they do it?’

**About every job depends on tourism.**

When the first Polynesian settlers arrived at Rapa Nui (Easter Island), there were probably only a few dozen of them. The island lies 2,150 miles west of South America and 1,300 miles east of its nearest inhabited neighbour, Pitcairn. Nowadays 12 flights arrive every

# Retrieval

# Ask people to find information

### **The Aral Sea is dry.**

October 2014: Satellite images from this week show that the eastern part of the Aral Sea is completely dry. 'It is likely the first time it has completely dried up in 600 years,' said expert Philip Micklin. The Aral Sea once had an area of 67,300 square kilometres. Two of Central Asia's biggest rivers, the Amu Darya and the Syr Darya, flowed into the Sea. This means that it was actually a freshwater lake, not a seawater lake. But it was so big it was called a sea. The Aral Sea used to be a busy place. Almost 20 percent of the Soviet Union's fish came from here and 40,000 people used to work near the lake. As the lake dried up, it separated into different parts. The eastern part nearly dried up in 2009 but it recovered in 2010 after some rain. Now, it's completely dry.

### **No fish can live.**

June 2015: Yusup Kamalov is a scientist from Uzbekistan. I am standing with him looking at a huge desert. Except that it's not like any other desert – fifty years ago the southern shore of the Aral Sea was right where we stand. Now it is 80 kilometres away to the northwest. We set off to drive to the water. On the way, we pass oil and natural gas rigs standing on the sand. According to Kamalov, each year there are a few more. 'Can you imagine,' he says, 'that 40 years ago the water was 30 metres deep right here.' We finally arrive at the edge of the lake, which is so salty that no fish can live in it. 'This is what the end of the world looks like,' says Kamalov.

---

Try to summarize the text above in three sentences.

Enter Summary Here

# **Comprehension**

# **Ask people to summarize the text**



# Retrieval and Retention

	<b>Accuracy</b>	<b>Time Taken</b>
Baseline (no titles)	0	0
Human-written	-0.01	-2.2 secs
Generated	-0.01	<b>-27.1 secs (p &lt; 0.01)</b>

# Comprehension

	Readability	Relevance	Length	Time Taken
Baseline (no titles)	4.66 ± 0.65	4.11 ± 0.86	0	0
Human-written	4.55 ± 0.76	4.09 ± 0.95	<b>+8.6 words*</b>	<b>-20.9 secs*</b>
Generated	4.52 ± 0.72	4.12 ± 1.02	<b>+5.3 words*</b>	<b>-2.6 secs*</b>

## **Section titles help with text comprehension**

The **type of title** influences what is remembered about a text

Extractive (Generated) titles: **Fact retention** is easier

Human-written titles: The **overall story is easier** to understand

# Conclusion

- We introduced a data-efficient title-generation pipeline
- We found that the SCRF-based compression outperforms S2S models in low-data settings
- We developed an evaluation framework and confirmed the positive effect of titles on text comprehension

But....

- The deletion-based compression only works for languages that retain grammaticality, even English has problems at times
- The efficacy of the low-resource model in an interface is still unknown

# Generated Section Titles Improve Text Comprehension

Sebastian Gehrmann, Steven Layne, Franck Dernoncourt



@SebGehr



gehrmann@seas.harvard.edu

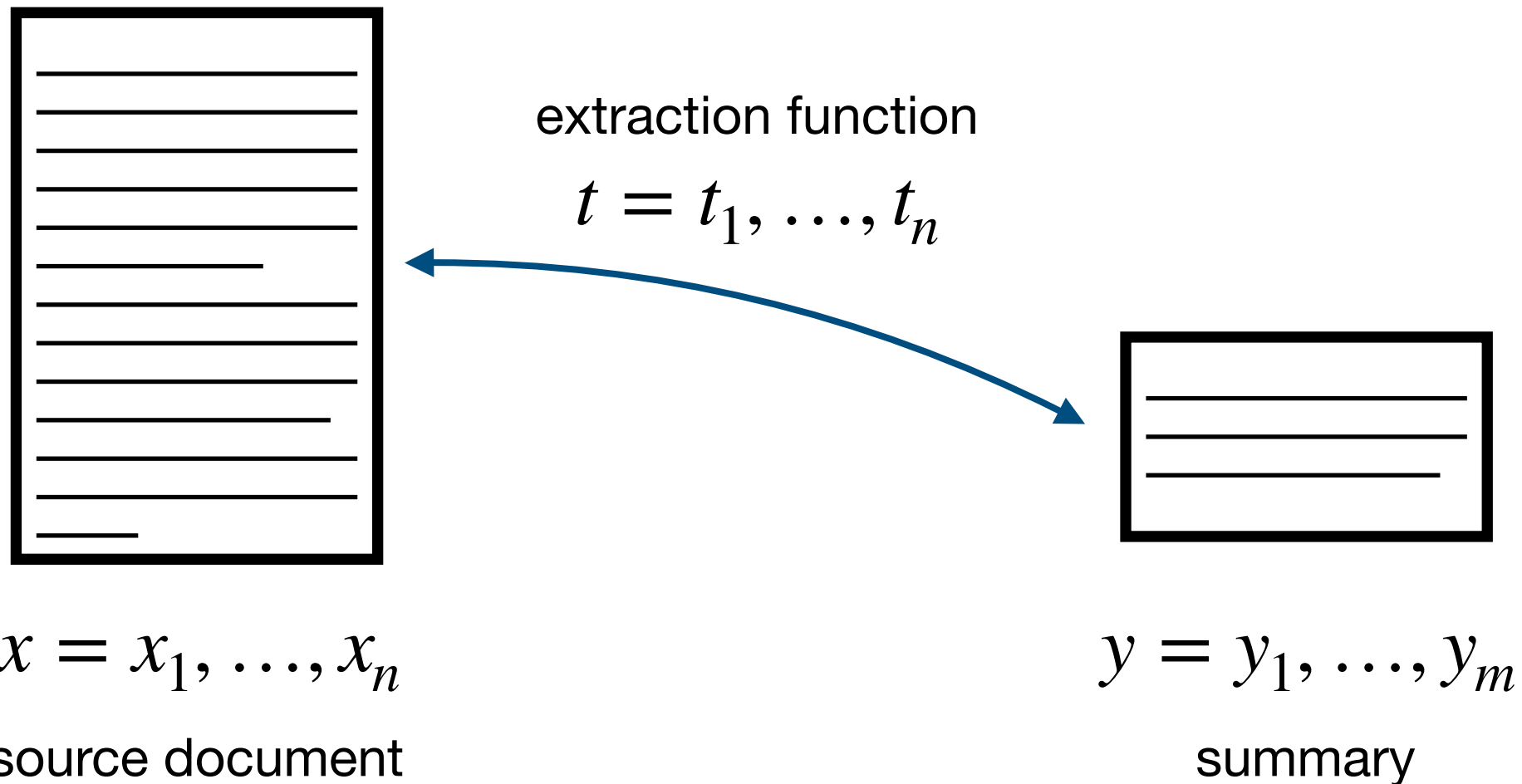


**Harvard** John A. Paulson  
**School of Engineering**  
and Applied Sciences



Adobe Research

# Sentence Selection through Word-Level Extractive Summarization



**Objective:** Learn  $\log p(t|x) = \sum_{i=1}^n \log p(t_i|x)$

**Selection:** Pick  $\text{saliency}(sent) = \frac{1}{|sent|} \sum_{i=1}^{|sent|} p(t_i|sent)$